# Cyberinfrastructure to support Real-time, End-to-End Regional Forecasting

by

Tom Baltzer, Doug Lindholm, Mohan Ramamurthy and Ben Dominico

## 1 INTRODUCTION

The past decade has seen a veritable increase in the number of local or regional modeling activities at universities and National Weather Service (NWS) Forecast Offices (WFOs). By a recent count, approximately 30 universities (http://rain.mmm.ucar.edu/mm5/pages/sites.html) and 50% of the WFOs (Rozumalski, COMET/UCAR, personal communication) are engaged in quasi-operational, near real-time local numerical weather prediction (NWP) efforts.

Three important factors have contributed to the proliferation of local NWP efforts: a) the availability of powerful but relatively inexpensive workstations based on commodity microprocessors and open source operating systems like Linux; b) availability of community models like MM5 and WRF; and c) easy, real-time access to data for generating initial and boundary conditions for model predictions.

Not only has the number of distributed local modeling activities grown, but so has the scope of those efforts. For instance, some universities like University of Illinois at Urbana-Champaign (UIUC) and University of Washington have set up mesoscale ensemble prediction systems using multiple modeling systems, different parameterization schemes, and/or multiple forecasts from a single model using different initial conditions.

The reasons for engaging in this challenging work are many fold. First and foremost, the demand for information from localized, mesoscale model predictions is an important driver for many of these activities. For example, more accurate specification, especially with regard to timing and location, of high-impact but poorly forecast sensible weather elements such as precipitation, diurnal temperature fluctuations, clouds and winds is a key motivation in a survey of forecasters in the WFOs (McQueen and Hirschberg, 2001). To that end, universities are increasingly collaborating with nearby WFOs and routinely sharing data from their local predictions. Notable efforts include those by the Pennsylvania State University (Warner and Seaman 1990), the University of Utah (Horel and Gibson 1994), the University of Washington (Mass et al., 2003), and the University of Oklahoma (Lazarus et al., 2002). These mesoscale forecasts, which augment guidance from operational forecasts from the National Centers for Environmental Prediction, are usually tailored to better predict local mesoscale circulations (e.g., terrain and lake-induced circulations and coastal surges) affecting their area. The ability to understand the impact of the various models' physics schemes by adjusting the parameterizations of the runs as well as to better understand the regional influences and sensitivities are also big motivations.

In a few instances, these local predictions take advantage of local observational assets by incorporating observations from nearby mesonets, although a vast majority of local real-time modeling sites continue to perform a so-called "cold start" initialization without any local data assimilation. In systems employing a cold-start approach, the local models are initialized simply from gridded analyses and forecasts provided by operational prediction centers.

Figure 1 shows an example of the some visualization from one of the local predictions from UIUC. The 3D rendering of the water vapor isosurface from a recent WRF 36-hour forecast, (10/29/04 00Z), nicely illustrates the relationship in a comma-cloud pattern between the upper-level jetstream-level winds (shown as contours of U-component of the wind near the tropopause) and the dry slot/moisture distribution within a mature mid-latitude cyclone.

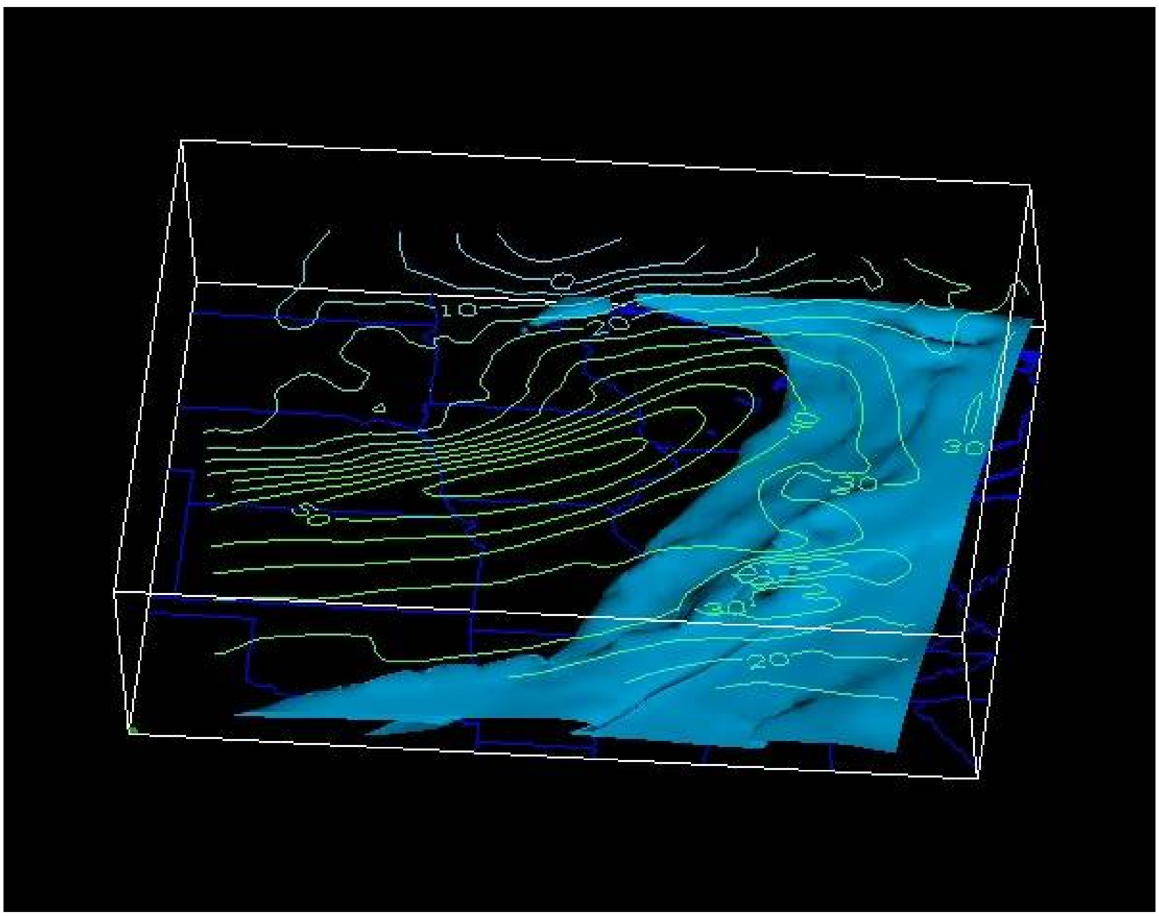


Figure 1: Example UIUC WRF data

## 2  How is it Being Done Today?

In order to perform local NWP, one must have timely access to current conditions to initialize the model and provide the appropriate boundary conditions.  Quite frequently this is accomplished at universities by obtaining initial and forecast grids from operational national forecast models like the Eta model or the Global Forecast System and using the interpolated zero hour analyses as initial conditions.

Members of the Unidata community obtain real-time observations and model grids through the Internet Data Distribution (IDD) system by running the Unidata Local Data Manager (LDM) software.  However, the model grids distributed via the IDD are at a coarser resolution than those used in the actual model prediction at NCEP, because the model grids ingested into the IDD are those provided by the NWS NOAAPort satellite broadcast.  Those universities wishing to use higher resolution model grids for their initial and boundary conditions use the Unidata CONDUIT (Cooperative Opportunity for NCEP Data Using IDD Technology) data stream, a special data stream set up by Unidata to facilitate real-time local NWP at universities.

The model grids provided by NCEP via either the IDD or CONDUIT are in GRIB (GRIdded Binary) format, a World Meteorological Organization standard format for archiving and exchanging gridded data.  GRIB is a binary format, and as such the data are packed to increase transmission and storage efficiency.  Being coded binary files, GRIB data are not readily useable without decoding by suitable software such as Unidata's *gribtonc*, conversion utility that transforms GRIB data into netCDF - a machine-independent format for representing geoscientific data.  This process is conceptualized in Figure 2.

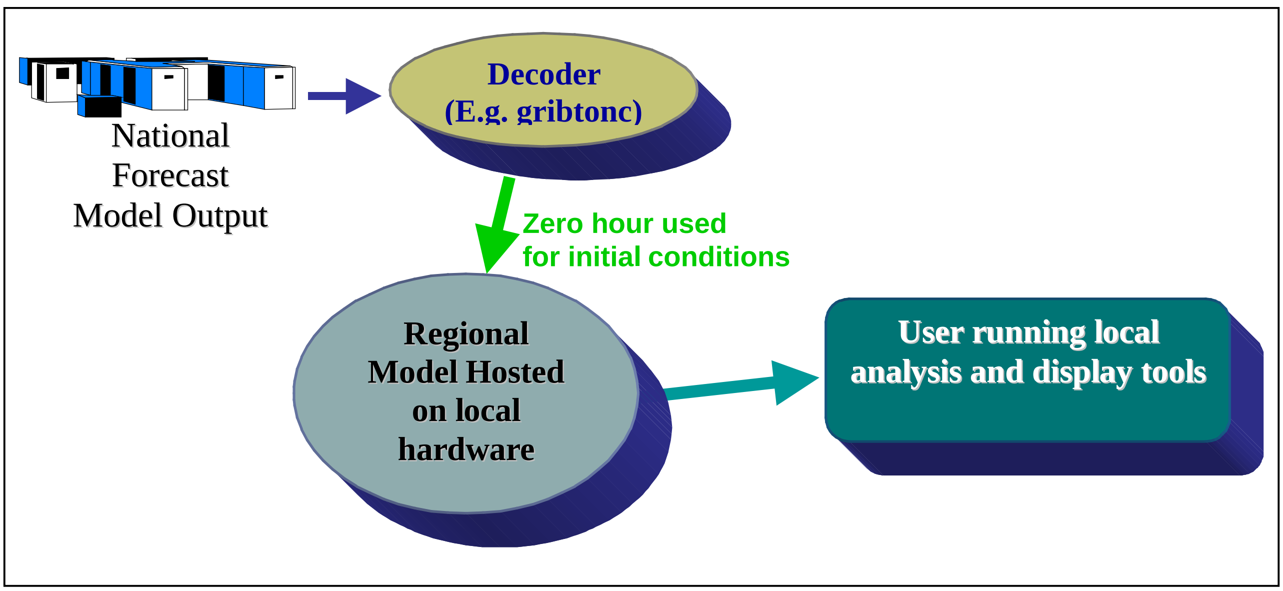

Figure 2: Initializing NWP with interpolated zero hour analyses ("cold start").

At operational centers like the NCEP, NWP models are initialized using a pre-forecast data assimilation cycle wherein current observations are combined using sophisticated algorithms with the background forecast fields from the previous cycle to generate the initial conditions. In fact, one of the motivations for universities to engage in local NWP activities is the desire to take advantage of local and regional real time observational assets for initializing the model. Inclusion of more local/regional observational data results in a more accurate specification of the initial state and perhaps improved predictions by the model. However, setting up a local data assimilation system is non-trivial, requiring not only access to local observations but also other complex software components for their inclusion in quality control and data analysis programs. As shown in Figure 3, an Assimilation system pulls together the various sources of data and puts them into a form that can be used to initialize the model.

In a recent survey conducted by Unidata, it was found that less than 10% of the real-time modeling efforts at universities and WFOs used a local data assimilation system to initialize their model.

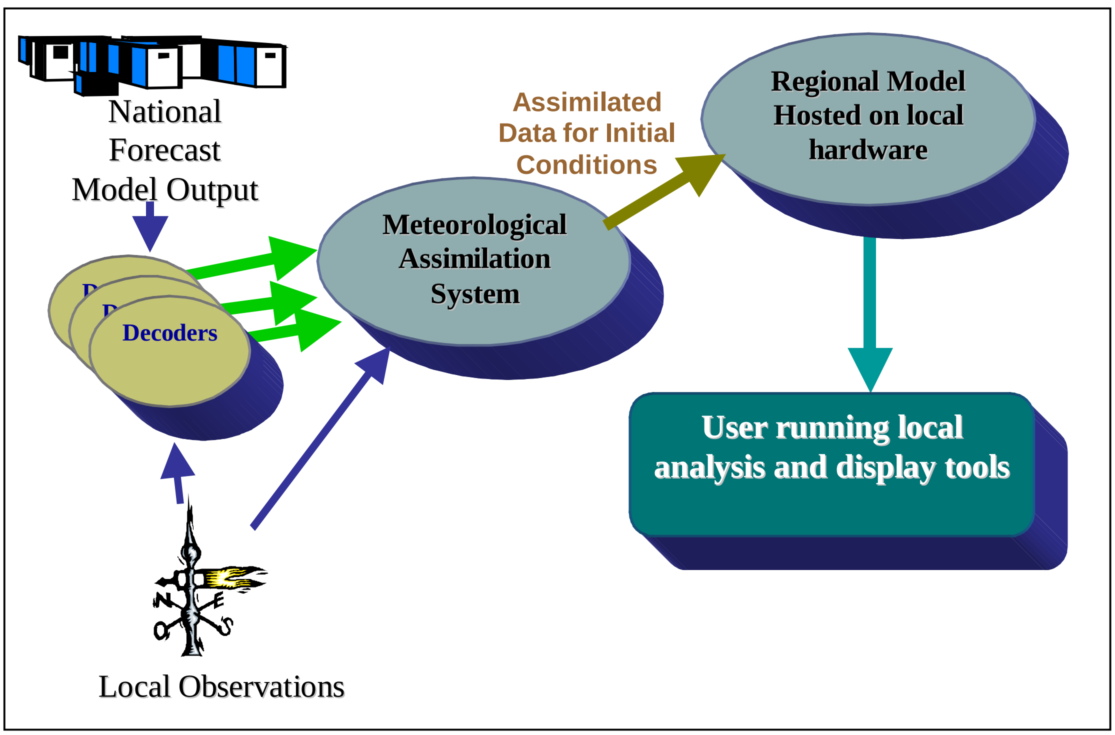


Figure 3: Regional Forecast with Improved Input

# 3 The LEAD Vision

The Linked Environments for Atmospheric Discovery (LEAD) is a National Science Foundation Large Information Technology Research (ITR) project that is creating an integrated, scalable cyberinfrastructure for mesoscale meteorology research and education. Further details on LEAD are provided in a companion paper at this conference by Droegemeier et al. (2005) and on the project website at http://lead.ou.edu/. LEAD is leveraging the technologies of Grid and Web Services and Service Oriented Architectures to provide the community with an easier way to perform regionalized mesoscale model runs.

According to Foster and Kesselman (1999), the Grid refers to an infrastructure that enables "*coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations*." Simply put, Grid computing facilitates the use and pooling of distributed computer and data resources to solve complex scientific problems and requires specialized middleware (the software glue that connects an application to the underlying operating system software or another application) to hide the complexity of creating and deploying grid applications. Because grid computing is still in its infancy, transparent, user-friendly middleware for connecting distributed computing environments to solve problems such as on-demand NWP does not yet exist. A service-oriented architecture (SOA) is essentially a collection of services. These services communicate with each other. The communication can involve either simple data passing or it could involve two or more services coordinating some activity. Much like Grid middleware, some means of connecting web services to each other is also needed.

The LEAD project is developing a Grid of high performance computers and storage to facilitate dynamic, on-demand NWP using the Weather Research and Forecast (WRF) model. The Grid will be enabled by the creation of a number of services for performing all the functions that go into generating these forecasts. A Web Portal front end is being developed that will allow a user to select from the available services in order to orchestrate a workflow for performing on-demand model predictions. This concept is illustrated in Figure 4.

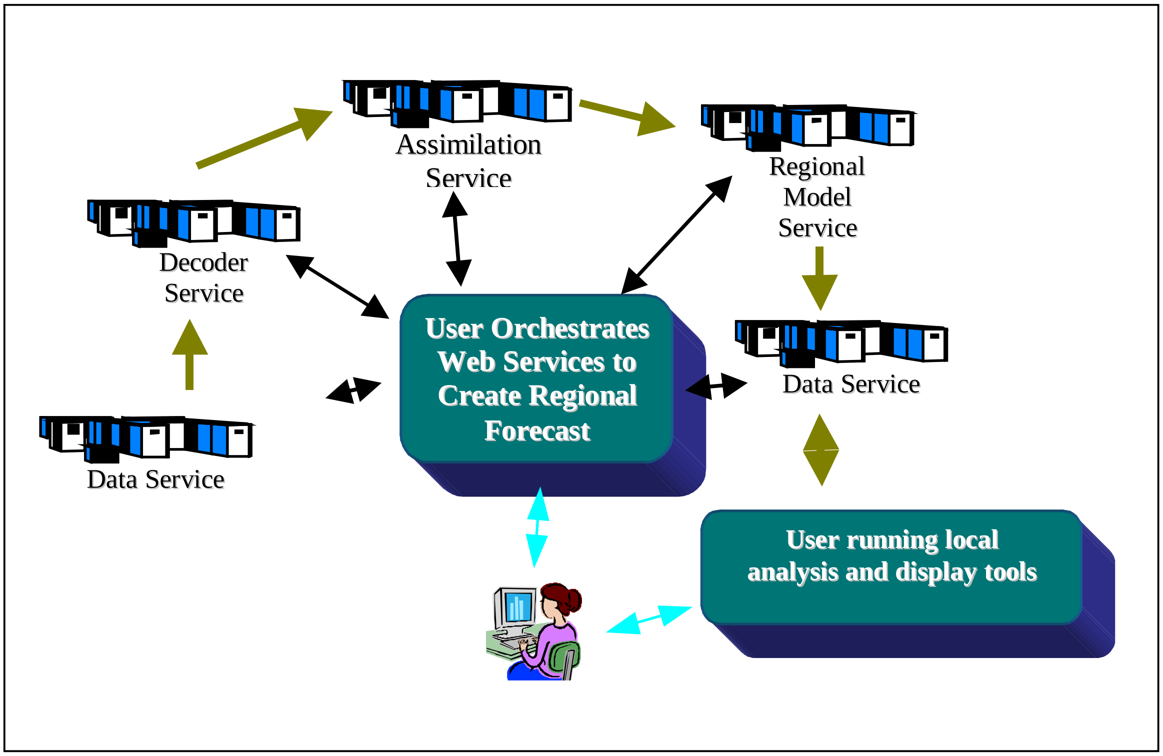


Figure 4: The LEAD Vision for Regional Forecasting

The idea is that the user would select a Data Service that would provide the initial condition information for their model run(s). They would then configure the workflow to take that data and send it to an appropriate decoder service so that the data is put into a form usable by the assimilation service. Next they would select the assimilation service and connect the decoded data to it. The data that is the result of the assimilation would then be connected to the Regional Model Service. And finally the user would select another (could be the same) Data Service for storing of the result. The portal interface will provide the user with the ability to adjust parameters of these services according to the capabilities of the underlying software (e.g. changing the physics scheme used by the Model).

In this way, a LEAD user will be able to perform regionalized mesoscale modeling without having to host a lot of complex software on their local systems. Further, by leveraging Grid technologies, they will be able to share their compute infrastructure with others in the community and vice-versa.

LEAD will be built not only upon the technologies of Grid and Web Services and Service Oriented Architectures; it will also rely heavily upon existing tools currently being used by the meteorological community for NWP functions. Table 1 provides a breakdown of the Tools upon which LEAD will be built, a description of their function, and a comment upon how the tool will be used within the LEAD architecture.

| Tool Name | Function | Comments |
|---|---|---|
| IDD – Internet Data Distribution | Network of institutions/software for delivery of near real time model and observational data to participating institutions | Provides model data for “cold start” NWP and observational data for data assimilation |
| LDM – Local Data Manger | Software for delivering IDD data to the Data Servers | Used to obtain model and observational data in real-time.. |
| OPeNDAP | OPeNDAP provides software (client/server) which makes local data accessible to remote clients regardless of local storage format. | Provides Data Server Functionality (sits behind Data Service) |
| ADDE - Abstract Data Distribution Environment | The ADDE allows your workstation to act as a client, efficiently accessing data from multiple McIDAS-X (data) servers. | Provides Data Server Functionality (sits behind Data Service) |
| THREDDS – THematic Real-time Environmental Distributed Data Services | Middleware that will facilitate the interoperability of data provider systems and the data analysis and display systems | Used in cataloging of near real-time data holdings and intermediate products. |
| NetCDF – Network Common Data Form | An interface, software library and self-describing data format that support the creation, access, and sharing of scientific data. | Data standard to allow interoperation of many Tools/Services |
| NetCDF Decoder Package | Software to convert from native format to NetCDF | Facilitates Tools/Service Interoperation |
| IDV - Integrated Data Viewer | A Java ™based software framework for analyzing and visualizing geoscience data. | Used to visualize observational data and model grids. |
| ADAS – ARPS Data Assimilation System | A 3-dimensional weather analysis program using model input for background fields coupled with real time observations for assimilation. | Used in the initialization of real-time predictions |
| ADaM - | A collection of tools for data mining and image processing. | Data Mining – LEAD will ultimately adapt processing behaviors according to |

| | | |
|---|---|---|
| | | current weather conditions. |
| WRF – Weather Research and Forecasting Model | A NWP system specifically designed for local and regional forecasting. | The mesoscale forecast model to be used by LEAD. |

Table 1: Tools to be used to Create LEAD

Ultimately, the vision of Service Oriented Architectures for Numerical Weather Prediction is to facilitate greater collaboration among researchers through the following mechanisms:

- Sharing of Results – at present, many institutions share results by putting up web pages displaying them. By instead cataloging and placing these results on well known data service providers' repositories it will facilitate sharing of model output with other researchers and educators in the community.
- Point of Presence Service Providers – Organizations that provide specialized research and operations capabilities will be able to provide their tools to the wider community by hosting web services to facilitate on-demand NWP.
- Orchestration at the desktop – While today access to LEAD will be through a portal interface, one can envision a time when a desktop application could be crafted to allow for orchestration of NWP workflow from the desktop interacting with Services put up by the Point of Presence Service Providers.

# 4 Conclusion

The desire for more accurate local forecasting is easily understood. Severe storms wreak havoc on society both economically and socially. Better prediction of these mesoscale events will lead to mitigation of property damage and saving of lives. The continuing trend toward inexpensive commodity microprocessor-based computing systems, ready access to sophisticated, community-supported mesoscale models, and increasing availability of real-time observations and model grids to initialize those models have resulted in the proliferation of real-time, quasi-operational NWP systems in the meteorological community, especially at Unidata universities and the WFOs.

Despite these advances, however, significant challenges still exist that prevent community members from exploiting the full potential of local NWP. Specifically, the current rigid frameworks, especially at the interfaces of the various components in a real-time modeling system, have been a significant barrier. Projects like LEAD and organizations like Unidata are working to develop tools to simplify the linkages between those individual components in order to better facilitate local NWP in the community.

## 5 Acknowledgements

The funding provided by the National Science Foundation under awards ATM 0331587 and ATM 0447781 is gratefully acknowledged.

## 6 References

Drogemeier, K.K., et al., 2005: Linked Environments for Atmospheric Discovery (LEAD): Architecture, Technology Roadmap and Deployment Strategy. *In press.*

Foster, I. and Kesselman, C, 1999: *The Grid: Blueprint for a New Computing Infrastructure.* Morgan Kaufmann.

Horel, J. D. and C. V. Gibson, C. V., 1994: Analysis and simulation of a winter storm over Utah. *Weather and Forecasting.*, **9**, 479-494.

Lazarus, S., C. Ciliberti, J. Horel, K. Brewster, 2002: Near-real-time Applications of a Mesoscale Analysis System to Complex Terrain. *Weather and Forecasting*, **17**, 971-1000.

Mass, C.F., et al., Regional Environmental Prediction over the Pacific Northwest. *Bull. Am. Meterol. Soc.*, **84**, 1353-1366.

McQueen, J., and P. Hirschberg, 2001: National Weather Service Local-scale Numerical Weather prediction Planning Workshop, December 12-13, 2000 at the NWS/Office of Science and Technology in Silver Spring, MD.

Warner, T. and N. Seaman, 1990: A real-time, mesoscale numerical weather prediction system used for research, teaching, and public service at Penn State Univ. *Bull. Am. Meterol. Soc.*, **71**, 792-805.